

Open Research Online

The Open University's repository of research publications
and other research outputs

Storyscope: Supporting the authoring and reading of museum stories using online data sources

Conference or Workshop Item

How to cite:

Mulholland, Paul; Wolff, Annika and Kilfeather, Eoin (2015). Storyscope: Supporting the authoring and reading of museum stories using online data sources. In: WebSci '15, ACM Press.

For guidance on citations see [FAQs](#).

© 2015 ACM

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1145/2786451.2786462>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Storyscope: Supporting the authoring and reading of museum stories using online data sources

Paul Mulholland
Knowledge Media Institute
The Open University
Milton Keynes, MK7 6AA, UK
+44 (0)1908 654506
p.mulholland@open.ac.uk

Annika Wolff
Computing and Communications
The Open University
Milton Keynes, MK7 6AA, UK
+44 (0)1908 274066
a.l.wolff@open.ac.uk

Eoin Kilfeather
Digital Media Centre
Dublin Institute of Technology
Aungier Street, Dublin, Ireland
+353 (0) 1 402 3272
eoin.kilfeather@dit.ie

ABSTRACT

Museum staff tell stories to assist visitor interpretation of artworks. Visitors also tell their own stories to articulate their understanding and opinion of artworks. Additional knowledge about the concepts mentioned or tagged in these stories can be found from online data sources. These could be used to assist reader interpretation or author development of stories. However, the potentially vast network of heterogeneous knowledge that can be created around the tags or annotations of a story could be bewildering for the story reader or author. Here we present Storyscope, a test-bed environment for the authoring, reading and semantic annotation of museum stories. The integration of online knowledge within the task of story authoring or interpretation is facilitated by mapping the available knowledge to a set of facts and simple events related to each story annotation. Narrative principles of theme and setting are used to discover and highlight aspects of the knowledge of potential value to the author or reader. Preliminary studies indicate the potential of the approach for providing a form of semantic navigation across stories and concepts having a better cognitive fit to story related tasks than existing forms of navigation.

Categories and Subject Descriptors

H.3.3 [Information Storage and retrieval]: Information Search and Retrieval; H.3.7 [Information Storage And Retrieval]: Digital Libraries—User issues; H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces—Theory and models

General Terms

Human Factors, Theory

Keywords

Museums, digital storytelling, story themes, story settings, events

1. INTRODUCTION

Museum objects are often accompanied by stories. The object may be displayed in a physical or online space alongside

the story. Alternatively, the object may be physically present with the accompanying story provided by a mobile device or tour guide. Stories are told by museum professionals to create a context and support interpretation of the object. These stories, used in catalogues, websites, visitor guides and wall panels, often take the form of a relatively short text supplying additional context to one or more museum objects by, for example, revealing more about the artist, the time in which the artwork was created or what is depicted in the artwork. Museum visitors also tell stories of their own [1]. Members of the public may tell stories to link objects to their everyday experience. Students may tell stories to express a viewpoint or demonstrate their understanding.

Online data sources such as Freebase [2], Jago [3] and Factforge [4] can offer additional knowledge associated with the concepts mentioned in the stories. The author could make use of the additional knowledge when considering how to develop the story. Equally, the reader could use additional knowledge to assist their interpretation and uncover implicit relationships between story concepts.

Linked Data browsers such as Sig.ma [5] and Sindice [6] can be used to navigate related knowledge, starting from concepts mentioned in the story, for example navigating from an artist to an artwork they created. One advantage of Linked Data browsers is that they can be used irrespective of the user's domain of interest. However, this style of navigation can be some distance away from the tasks or goals of the user.

Knowledge related to an annotation can also be gathered into summaries or visualizations that assist the user in building an interpretation across a larger number of concepts. For example, data related to an artist could be gathered into an artist's biography [7] or a set of data points could be presented on a chart. However, these representations are appropriate for particular types of task and data and may not align, for example, with the goal of a reader or author, or the domain in which they are working.

Providing readers and authors of stories with appropriate ways of navigating associated knowledge requires an understanding of the nature of stories and the cognitive activities undertaken when creating or reading a story. Narrative theory offers insights into how stories are structured and the processes by which they are created and understood. The aim of our work is to draw on narrative theory to develop ways of linking stories and their surrounding knowledge that has a better cognitive fit with the types of activity the reader or author may wish to undertake.

We propose to employ the narrative concepts of setting and theme to provide a focus and abstraction for how the potentially

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

WebSci '15, June 28 - July 01, 2015, Oxford, United Kingdom
© 2015 ACM. ISBN 978-1-4503-3672-7/15/06...\$15.00
DOI: <http://dx.doi.org/10.1145/2786451.2786462>

large knowledge space around the story is explored, in which themes are key concepts (e.g. people, objects, art movements) of the story and settings are times and places at which events in the story occurred. This approach was inspired by some of our earlier work using museum stories, manually annotated according to their constituent events, to understand what the story reader saw as important [8].

The rest of this paper is structured as follows. The next section looks at the nature of narrative construction and interpretation and its relationship to previous work on the application of semantic web technology in museum contexts. Section 3 describes Storyscope, a test-bed system for the management of museum stories and describes how the narrative principles of theme and setting can be used to navigate between stories and associated knowledge. This is followed by a description of the Storyscope ontology that can be used to reason over stories across multiple Storyscope sites. Finally, a series of evaluations are described, investigating the utility of theme and setting in supporting story development and interpretation.

2. RELATED WORK

Theories of narrative construction, structure and interpretation can shed light on the types of activity that the creators or readers of narratives undertake. The theory of narrative inquiry [9] offers a perspective on how narratives are constructed from historical evidence. This is particularly relevant to museum storytelling where the curatorial process involves the assembly of evidence and its presentation in the form of narratives. The process of narrative inquiry involves four main stages. First, historical events are organized in time to form a chronicle. Second, the chronicle is divided into separate thematic strands focusing, for example, on different characters, locations, objects or types of event. This can bring together events that may be potentially related during the process of inquiry. This categorization of events into thematic strands is termed a storyline. Third, events within the storyline are emplotted, imposing relationships between them such as causal relationships between events. Finally, the emplotted story is formed into a narrative where the events and their relationships are organized and presented, for example, as a text or film.

Narrative theory can also shed light on the structure of the resulting narrative and its interpretation by readers. Gervás and León [10] describe the different types of inference that a reader can make when interpreting a narrative. One of these concerns theme, in which the reader infers what the story is about. Narrative theme may relate to the thematic strands of the underlying storyline, such as key people within the story. Theme also relates to the moral or point of the story [10] and is also characterized as the central concepts that bring coherence to the overall narrative [8, 11].

Readers also make a series of inferences related to the temporal or causal relationships within the story, for example, identifying why a person acted in a particular way at a particular time. Causal and temporal inferences require an understanding of the settings of a story. A setting is a time and location at which events of the story occurred [8]. Determining settings facilitates the identification of potential causal relationships by understanding whether one event preceded another or whether two people met in a particular time and location.

In this work we adopt setting and theme as two narrative constructs that could be used to support navigation across stories

and associated concepts in a way that is nearer to the cognitive activities of the story author or reader but still independent of the particular domain of the story.

The nature and organization of events is therefore important to how stories are built and understood. From a semantic web perspective, event-based representations are commonly used to represent and reason about cultural heritage knowledge. The CIDOC CRM ontology [12], developed for use in the cultural heritage sector, affords an event-based representation of knowledge in which, for example, the dimensions of an object can be represented as a measurement event. This allows representation of changes in the dimensions of an object over time. A number of event ontologies have also been developed such as LODE [13] and SEM [14]. Unlike CIDOC CRM, these essentially streamline event properties to variations of who, what, where and when, in order to have a standardized schema across a range of event types.

Some work has looked at how an event-based representation can support navigation through cultural knowledge. Hyvönen et al. [15] used an event-based approach in the development of a gazetteer about the First World War from Linked Data sources. Work on digital hermeneutics [16] has investigated how event properties such as times and locations can be used to navigate across an event space. In our work we investigate how the narrative constructs of theme and setting can be used in combination with an event-based approach to provide a new form of navigation across stories and their related concepts.

3. STORYSCOPE

Storyscope was developed as a test-bed environment for creating and publishing museum stories. Storyscope was built using the Drupal Content Management System [17]. A Storyscope site is organized into a series of dossiers, a metaphorical reference to the cardboard dossier that a curator may use to assemble materials when putting together an exhibition. A dossier is used to collect together a set of stories and associated media around a particular topic. Each story contains text, an image gallery and semantic annotations (see figure 1). Annotations are added to a story using a variant of the Freebase suggest widget. Alchemy API [18] is used to suggest additional annotations from the text of the story. When an annotation is associated with a story, external knowledge related to the annotation is retrieved and cached locally within the Storyscope site. In the current version of Storyscope, data is retrieved and cached from Freebase. This will later be extended to other textual and Linked Data sources.

In the following subsection we describe how the retrieved knowledge associated with the story annotation (which in Storyscope we refer to as ‘story tag’ and which is the term that will be used from now on) is organized and presented in order to assist the story author or reader. In section 3.1 we describe how knowledge associated with a story tag is organized and presented as a set of facts and events that comprise a *concept space* for the story tag. In section 3.2 we describe how themes are extracted from the concept space in order to highlight key people (e.g. artists), objects (e.g. artworks), ideas (e.g. artistic styles), etc. of potential interest to the reader or author. In section 3.3 we describe how key settings (i.e. times and places) are extracted from the concept space and can be used for further navigation.

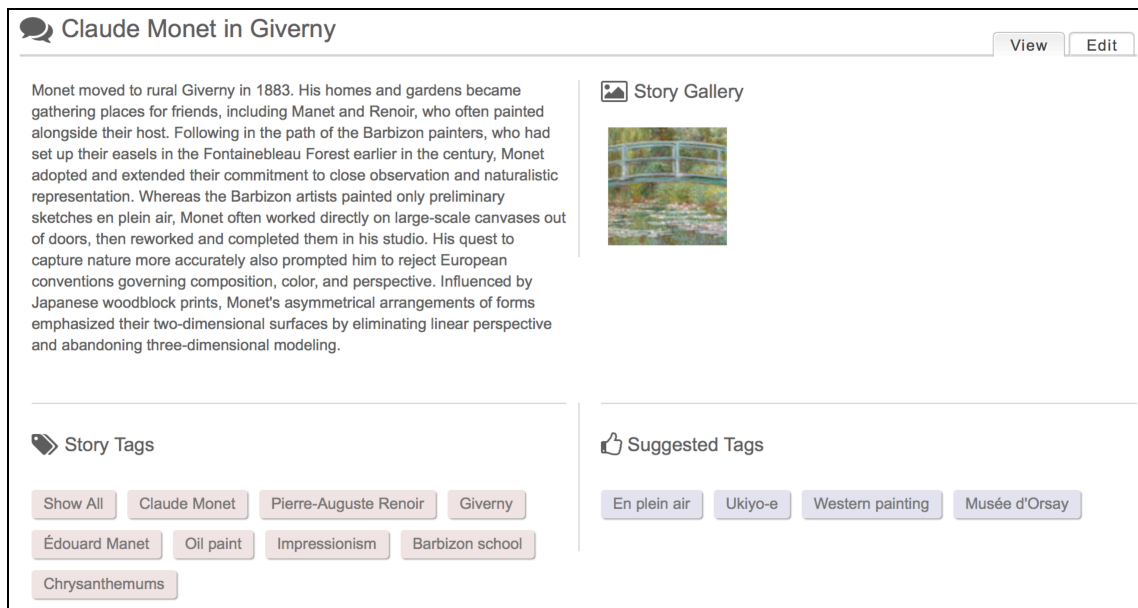


Figure 1. A story with image gallery, tags and additional suggested tags from the text.

3.1 Representing facts and events

Storyscope stories are annotated with story tags. Knowledge related to the tags is retrieved from an external source, currently Freebase. If the tag relates to a person, then data such as place and date of birth and associated art movements is represented as factual attributes about the person. Data such as places lived, objects created, books authored and education history is represented as events in which the person was involved. This distinction between what is an event and what is a fact, while by no means fixed, is based on an observed preference of museum professionals during our work. It also aligns with the observation made by Makela et al. [19] who found that library indexers prefer to model birth and death as properties of a person rather than events in which they are involved.

The facts and events of a story tag are retrieved from Freebase using the topic and MQL APIs. The topic API is used to retrieve the name, description and associated image of the story tag. The topic API is also used to retrieve the notable types of a Freebase topic (e.g. whether a person is primarily known as an artist, author, actor, etc.). The Freebase MQL API is used first to retrieve the types of a Freebase topic (e.g. location, person, artwork), then a series of further MQL queries are run depending on the types returned. For example, if a topic has the type “people/person” then a query is used to retrieve date of birth, date of death and education history. Birth and death are then internally represented as facts of the story tag. Education history is

represented as events associated with the person. If a topic had the type “/visual_art/art_subject” (i.e. has been a subject of one or more artworks), then the associated artworks are represented internally as artwork creation events associated with the subject. Currently, 34 different MQL queries are used to retrieve information about 23 different topic types.

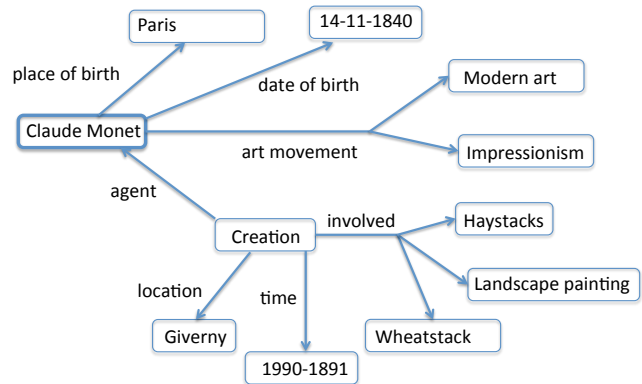


Figure 2. Representing facts and events related to Claude Monet.

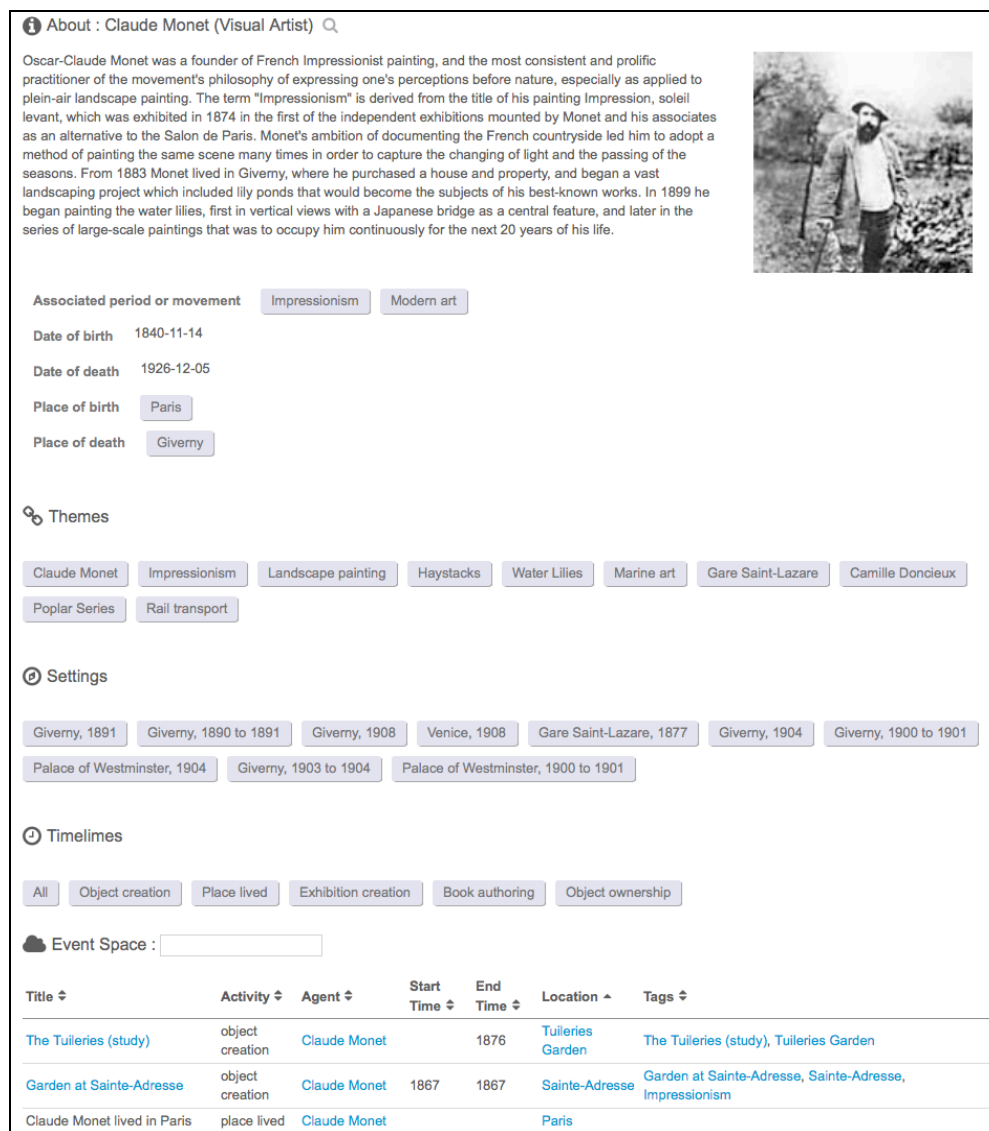


Figure 3. Facts and events related to Claude Monet presented in Storyscope.

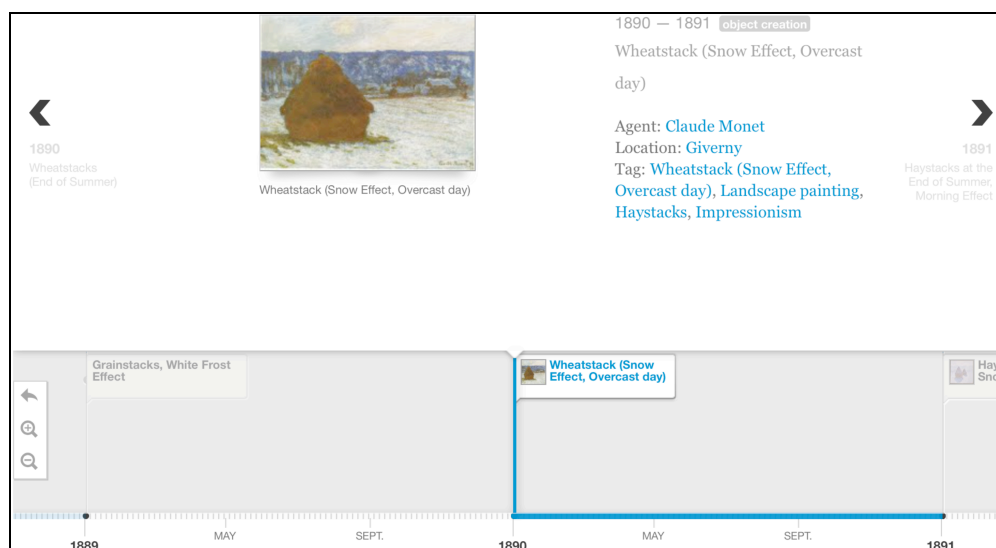


Figure 4. Timeline of events.

Events are represented using a simple event schema aligned with the LODE [13] ontology. The event properties are type, agent, location, time and involved. The *involved* property relates any other concepts associated with the event. Figure 2 shows a selection of the concept space related to Claude Monet, represented using fact and events. Place and date of birth and associated art movements are represented as facts. The creation of the artwork Wheatstack from the Haystacks series of paintings is represented as an event. Selecting a story tag within the Storyscope interface (see figure 1) brings more information specific to it into focus. Selecting the "Show all" tag produces a combined concept space (of facts and events) across all tags of the story.

Figure 3 shows how facts and events of a story tag (in this case Claude Monet) are represented in the Storyscope interface (for reasons of space, only a small number of the events are shown). The events are organized into a table. The set of events can be filtered and sorted according to its properties. A title is generated for each event based on its properties. Following user feedback, for increased comprehensibility within the Storyscope interface, time is split into start and end times of the event and the involved property is labeled as tags of the event. The events of an individual story tag or all tags can be visualized on a timeline (figure 4).

The schema used to describe events in Storyscope was iteratively designed with museum staff. An earlier iteration of Storyscope used different types of events each with specialized properties. So, for example, an event to represent the purchase of an artwork had specialized properties for representing the buyer, seller and price. This is a style of modeling that can be done using the CIDOC CRM ontology, in which specializations of a generic event class can be defined and used. In our testing of Storyscope this was found to be too complex by users and replaced with a who, what, where and when approach to representing events.

As it stands, the event space can be thought of as a chronicle of associated events. Focusing on the events of an individual story tag, or filtering the event space essentially allows the user to explore different storylines associated with the story. However, the combined events of the story, or even those of individual story tags, can produce a relatively large event space of over 100 events that may be difficult for the author or reader to use in the development or interpretation of the story. The narrative principles of theme and setting are used to provide an entry point into this potentially large space of associated knowledge and support a form of navigation intended to provide a closer fit to the activities of the author or reader.

3.2 Identifying themes of a concept space

As described earlier, themes are central concepts that bind together the other elements of the story. The story already has a candidate set of themes in the form of its tags. Additional concepts (e.g. people, objects) contained in the concept space of the story are evaluated as candidate themes, in terms of how they bind together the tags of the story. The concepts are scored in terms of:

- Coverage - How many story tags they are associated with either as facts (such as art movement of an artist) or through co-occurring in an event with the story tag.

- Frequency - How many times the concept appears in the concept space either in a fact or event associated with the story tag.

The candidate themes of the concept space are then sorted primarily in terms of coverage and secondarily in terms of frequency. Theme concepts are generated from the *agent* and *involved* properties of the events and any fact properties (such as associated art movement) that are specified as relevant to theme in the Storyscope configuration panel. A location or even a time may appear as a theme if it is described as being involved in the event rather (or as well as) a location or time property of the event. For example, a location may be described as the subject of a painting but not necessarily the location at which the painting was created.

If a story is associated with the tags Claude Monet and Paul Cézanne (see figure 5) those two tags would be associated with a large number of facts and events (just four events are shown for illustration). If as in the figure both were involved in education events related to the same art school (Académie Suisse), then Académie Suisse would have a higher theme rating than the other concepts that may be more frequent in the combined concept space but only associated with one of the story tags.

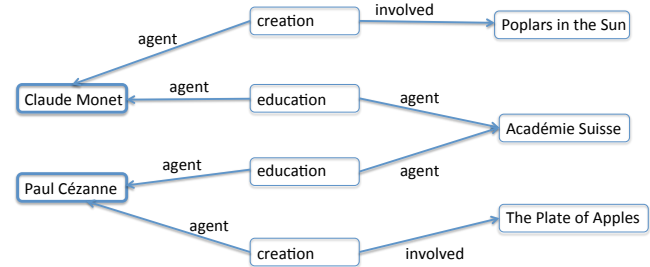


Figure 5 Deriving themes from the story tags Claude Monet and Paul Cézanne.

In the future we wish to contrast this approach with alternative methods for determining theme drawing further on narrative theory. This may involve coverage and frequency having tunable weightings rather than being primary and secondary means of ordering. Different event properties could also have customized weightings, for example agent(s) of events being more important than other event attributes. Additionally, a graph-based approach, such as PageRank [20], could be used to take into account indirect relationships between tags and event attributes. For example, an event attribute (such as a person) could have no direct connections to the story tags via a single event but be two event steps away from all of them.

3.3 Identifying settings of a concept space

Similarly, settings are generated for the space of events associated with a single story tag, or all tags of a story. The candidate settings of an event space are all the time-location pairs that can be derived from the time and location properties of the events. A setting may include a time point (represented in the Storyscope interface as only a start or end time) or a time span (for events that have both a start and end time). Candidate settings are ranked using a similar approach to theme ordering. Settings are primarily ordered according to coverage, defined as the number of story tags associated with events featuring that particular setting. Frequency is again used as the secondary ordering principle. For example, if a story has tags related to the

two paintings, *Poplars in the Sun* and *The Four Trees* (see figure 6), the setting Giverny 1891 would have a higher rank than candidate settings more frequent in the space of events but associated with fewer story tags.

Settings are calculated in two passes through the data. In the first pass, the frequency and coverage of each unique setting is calculated. On the second pass, settings of the same location with overlapping time specifications are merged. For example, the settings Paris 1990-1902 and Paris 1901-04 would be merged as Paris 1900-04. The new setting aggregates the coverage and frequency scores of the settings from which it was merged.

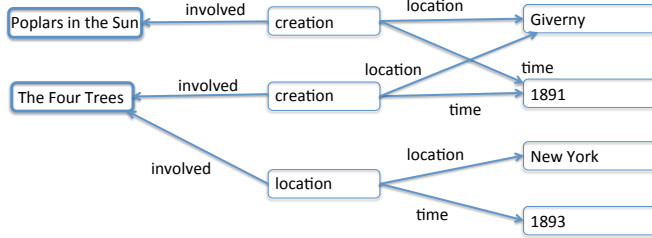


Figure 6. Settings derived from the story tags of two artworks.

Themes and settings are displayed in the interface as shown in the middle of figure 2. Selecting a theme produces the space of facts and events related to that concept. Selecting a setting produces the set of events matching that setting. Additional events are retrieved that match or overlap in time with the setting and have a location that either matches or contains the setting location. This brings in events that provide a broader historical context for the setting. For example, if the setting refers to a city or location within a city, larger scale social and political events affecting the country or wider region at that time, such as elections and wars, are included as historical context.

Themes and settings therefore provide an entry point into the data found in the fact and event space by identifying key concepts that also directly support the reader or author in pursuing lines of inquiry. An identified theme, such as Claude Monet, can lead the reader or author to pursue a thematic question such as “What else did Claude Monet do?” A setting such as “Giverny 1891” can lead the author or reader to pursue a setting-related inquiry such as “What else was happening in Giverny in 1891?” Thematic and setting-related inquiries could help the reader to interpret and understand the context of the story as well helping authors to extend the story in new coherent directions.

Themes and settings are generated and displayed not only on the level of the story but also for dossiers and the whole Storyscope site. In these cases, themes and settings are ranked in terms of the number of stories they interconnect, secondarily in terms of story tag coverage and third, in terms of frequency. Themes and settings provide an alternative entry point to story content within a Storyscope site. For example, the reader can select a theme of the site or one of its dossiers, explore related facts and events and access stories associated with the theme. Navigation can therefore lead the user both from stories to concepts and from concepts back to stories.

4. STORYSCOPE ONTOLOGY

The Storyscope ontology [21] was developed to describe the structure and content of a Storyscope site. An RDF dump can be generated of any Storyscope site according to the ontology. An overview of the ontology is shown in figure 7. The overall

structure of the Storyscope site is described using a specialization of the SIOC ontology [22]. Storyscope sites, dossiers and stories are specializations of the SIOC concepts of space, container and item respectively.

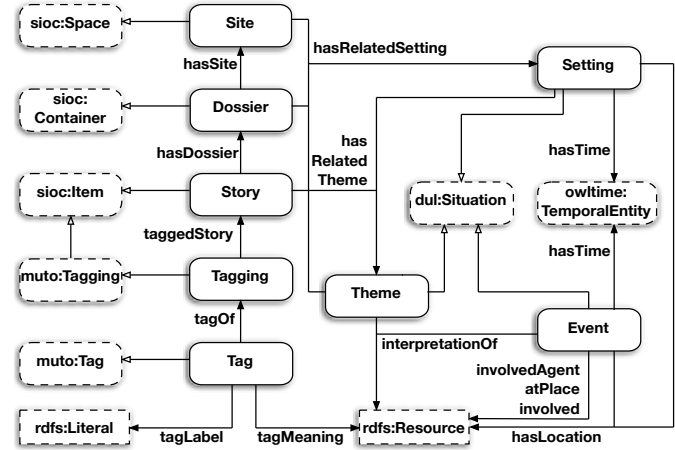


Figure 7. Overview of Storyscope ontology.

The relationship between semantic tags and stories is described using a specialization of the MUTO ontology [23]. A set of tags is associated with a story via a tagging. Even though the current Storyscope system does not allow for multiple or alternative taggings of the same story, this separation of stories and tags is still useful for reasoning purposes. For example, if identical stories were published in separate Storyscope sites (or duplicated within an individual site) an owl:sameAs relationship could be declared between the stories. This would still preserve the alternative taggings of the duplicate stories rather than merging them into a single set of tags. The label and meaning of each tag is also represented using the properties tagLabel and tagMeaning.

The events derived from external sources (e.g. object creation, book authoring, etc.) and also the themes and settings are classified as situations according to the Descriptions and Situations ontology design pattern [24]. This allows representation of both the event and its properties (i.e. involvedAgent, atPlace, involved, hasLocation, hasTime) and the relationship to the external resource of which the event is a re-representation.

The ontology allows the comparison of descriptive statistics across sites, such as the number of dossiers, stories and tags. The ontology also allows for the comparison of the interpretative activities being carried out in each site. For example, it is possible to explore differences in stories associated with the same concept (such as a person or museum object), identifying how different communities variously interpret the work of an artist or a particular artwork. Similar analyses can be carried out over the themes and settings emerging from different sites and their relationship to the authored dossiers and stories.

5. EVALUATION

Three preliminary evaluations were carried out to identify the potential of theme and setting as useful abstractions when navigating across stories and their associated knowledge. The first was a small user study to investigate whether readers were able to make use of the concept space of each story tag when interpreting a story. A comparison was made to other sources containing similar information content. The second and third studies

investigated whether the methods of theme and setting identification respectively could be used to identify emerging themes and settings in a set of museum stories aimed at visitors. These studies are described in turn.

5.1 User Study

A small observational study was conducted to gain insight into how theme recommendation could support searching for different types of information in comparison to other information sources, namely web pages and Freebase. A task was devised around two artists, Dante Gabriel Rossetti and William Holman Hunt, who were both founders of the Pre-Raphaelite Brotherhood. The first question asked ‘what artistic movement were they linked to?’ in which participants should find that they were both founders of the Pre-Raphaelite Brotherhood. The second required participants to make a value judgement and asked ‘what important artworks did they create?’ The final question asked ‘what relationships can you find between them and other artists?’

Participants were either given access to i) Freebase pages about the two artists ii) the concept spaces generated for each artist, plus the multi-tag page that merged both artists, or iii) an artist biography for each taken from www.tate.org.uk. The Tate pages were selected because the information content and formatting was similar for both artists and was also similar to the information available in the Freebase and Storyscope conditions. Participants first undertook a practice task to ensure that they were familiar in navigating each type of resource. Participants were recorded using screen capture as they browsed and wrote down their answers. Their talk aloud protocols were recorded. There were two participants for each version (six in total). One participant was not recorded using screen capture, but their observational data was included.

For the first question we recorded the length of time it took for participants to find the answer from the source they were given. These preliminary findings suggest that the concept spaces facilitate finding information that links artists. Participants were significantly faster in finding the information about the Pre-Raphaelite Brotherhood and linking it to both artists than in other conditions (around 1.13 minutes on average, compared to between 2 -5 minutes in the other conditions). However, participants relied heavily on written text for making value judgements about important artworks. In the Freebase and Storyscope conditions several participants queried ‘how do I know which are important?’ before searching for written text that allowed them to make a judgement, such as ‘first major oil painting’. One participant in the Storyscope version merely copied the entire lists of works. For the third question, when describing the relationship between artists, without fail the participants used the terms available to describe the nature of the relationship, rather than discovering or choosing their own. In the web case, participants used terms directly from the written text such as ‘was a pupil of’ or ‘was friendly with’. In the Freebase condition participants listed the artists who ‘influenced’ or ‘were influenced by’ each artist. Across all conditions, participants also listed other artists who, according to their information source, were involved with the Pre-Raphaelite Brotherhood.

In summary, the concept space and associated themes appeared to facilitate discovering information that linked more than one artist, whereas a web-type resource made it easier to answer a question which required a value judgement to be made, so that participants could rely on the judgement of others who had written the text (participants did not appear to naturally want to

dig further and decide importance for themselves). Similarly when describing relationships, the information available heavily influenced what was selected and the terms used to describe the relationship, even though further searching of the resources could have provided further information that could have been used.

5.2 Predicting Related Themes of Museum Stories

As described earlier, the aim of suggested themes is to introduce concepts that the author could use to extend the story or the reader use to interpret and contextualize the story. If we assume that a story is written linearly, from beginning to end, the themes proposed from a tag early in the story could predict tags appearing later in the story. Even if the stories are not written in a linear fashion it can be expected that concepts (and therefore tags) introduced early in the story will lay the foundation for concepts introduced later in the story. One challenge is that the initial story introduction could potentially develop in a large number of coherent ways. For example, a story that starts by introducing an artist could move on to discuss the artist’s works, contemporaries, personal life or the large scale political and social influences in which they worked.

The stories related to artists and artworks from the late 17th to the early 20th Century. Stories intended for use by visitors and for publication in guidebooks are relatively short. The mean number of words per story was 133. Each story was tagged according to its Freebase annotations. The mean number of tags per story was 5.2. For this experiment we took a set of 32 stories aimed at visitors provided by the National Gallery of Ireland. The set of stories were similar to the artist biography stories from the user study in terms of their length and the number of tags with which they could be associated.

For each story, the concept space, and the themes derived from it, was generated from the first 1, 2, 3 and 4 story tags. In each case, the top 10 suggested themes were generated (as displayed in the Storyscope interface). The set of themes were generated in two ways. First, themes were produced, ranked primarily by coverage and secondarily by frequency, as described earlier. For comparison, themes were also ranked solely in terms of frequency. Themes were not also solely ranked in terms of coverage as this creates a number of ties due to multiple concepts being related to the same number of story tags in the concept space.

Table 1. Predicting future tags from the first 1, 2, 3 and 4 story tags.

Annotations	1	2	3	4
Stories	29	27	22	17
Remaining tags	4.72 (3.17)	4.0 (3.11)	3.68 (3.06)	3.47 (3.06)
Concept space	1.38 (1.70)	1.67 (1.61)	1.63 (1.26)	1.59 (1.77)
Coverage, frequency	0.62 (0.78)	0.56 (0.64)	0.55 (0.67)	0.53 (0.72)
Frequency	0.62 (0.78)	0.67 (0.78)	0.68 (0.78)	0.59 (0.87)
h0	0.30 (0.49)	0.23 (0.35)	0.18 (0.28)	0.13 (0.26)

For each method, a count was made of the number of future story tags appearing in the theme list. The two methods were compared against a null hypothesis (i.e. h_0) in which the themes were selected at random from the concept space. Table 1 shows the first 1, 2, 3, and 4 tags of each story, the number of stories remaining in the analysis at that point (i.e. some stories had fewer than four tags), the mean number of remaining tags, the mean number appearing in the overall concept space, and the mean number appearing in the top 10 themes according to coverage and frequency, frequency only, and the null hypothesis. Standard deviation is shown in brackets. The number of predicted tags for coverage and frequency and frequency only is identical for the first annotation as coverage and frequency is reduced to ordering by frequency for a single annotation as each concept links only to the sole story tag.

By comparing the mean number of remaining tags in the story (i.e. 4.72 after one annotation) with the mean number of remaining tags found in the concept space (1.38 after one annotation) it can be seen that the percentage of remaining annotations in the concept space gradually increases from 29.2% after one annotation to 45.8% after four annotations. This can be seen to reflect the expansion of the concept space with each story tag. Also, as the story evolves and becomes more detailed, subsequent annotations can be expected to be conceptually close and therefore more likely to co-occur in the same concept spaces.

Coverage and frequency was found to be a significantly better predictor than the null hypothesis for the first 2 ($t = 2.30$, $p < 0.05$), 3 ($t = 2.35$, $p < 0.05$) and 4 ($t = 2.16$, $p < 0.05$) story tags. Frequency alone was found to be significantly better than the null hypothesis for the first 2 ($t = 2.62$, $p < 0.05$), and 3 ($t = 2.83$, $p < 0.01$) story tags. No significant difference was found between coverage plus frequency and frequency alone.

5.3 Predicting Related Settings of Museum Stories

The aim of the suggested settings is to introduce the author or reader to settings that may be of use when extending or interpreting the story. This would suggest that the setting recommended from the story tags might predict settings explicitly mentioned in the story. For this experiment, any settings explicitly mentioned in a story such as Paris 1880-1881 were identified. The set of 34 stories used for theme prediction (section 5.2) contained 20 explicit settings across 15 stories. The tagging of stories, using the variant of the Freebase suggest widget, included locations explicitly mentioned in the story but not time specifications. The top 10 story settings were generated for each story that contained one or more explicitly mentioned settings.

As in the theme experiment, the top 10 settings were calculated using either coverage plus frequency or frequency only. For each story it was calculated how many proposed settings matched the explicit settings of the story. A proposed setting was considered to constitute a match to an explicit setting if the proposed setting had the same or sub-location of the explicit setting and an overlapping time specification, so for example, Paris 1880-1881 or Pont Neuf, Paris, 1881, would be considered as a match for Paris 1880-1882. A measure was also taken of how many proposed settings matched only the location or only the time of the explicit settings.

The two methods were compared against a null hypothesis, in which 10 candidate settings were selected at random from the event space of the story tags. This was done 10 times for each

story and a mean score calculated. Table 2 shows the number of proposed settings that match an explicit setting in the story. Results are given for ordering by coverage and frequency, frequency only, the null hypothesis, and the number of matching candidate settings across all events in the event space. Rank ordering by coverage and then frequency produces a higher number of complete matches than frequency ordering or the null hypothesis, although this is not statistically significant. Table 3 shows the number of explicit story settings that are matched by one of the proposed settings. Of the 20 explicit settings, 8 completely matched a candidate setting in the entire event space and 5 of these matched a setting selected according to coverage and frequency. Ordering by coverage and frequency produced more complete matches than frequency ordering or the null hypothesis though this is also not statistically significant.

Table 2. The number of proposed settings matching an explicit story setting.

	Coverage, frequency	Frequency	h_0	Event space
Time & location	6	2	2.6	22
Location only	36	31	35.8	133
Time only	6	5	6.5	40
Total	48	38	44.9	195

Table 3. The number of explicit story settings matching a proposed setting.

	Coverage, frequency	Frequency	h_0	Event space
Time & location	5	2	2	8
Location only	10	6	11	9
Time only	2	2	3	3
Total	3	10	4	0

5.4 Evaluation Summary

In summary, the small user evaluation provides evidence that readers are able to use concept spaces and themes to identify additional concepts that connect story tags. Separate evaluations using a pre-existing set of museum stories suggest that theme and setting could be of use to the reader and author in offering concepts of potential use when interpreting or extending the story. In terms of theme, the concept space associated with story tags contained a number of future tags. Themes ranked by coverage and frequency and frequency alone provided relatively good predictions of future tags. In terms of setting, the concept space contained a number of the explicitly mentioned settings and ordering by both coverage and frequency provided the best prediction of explicit settings, though this was not statistically significant. Future studies will look in more detail at how theme and setting can support the cognitive activities of story readers and authors.

The user study asked participants questions concerned with how the two artists were connected and the important artworks the two artists created. Evidence from the theme prediction study provides further evidence that Storyscope can assist authors or readers in answering this type of question. Themes are identified that do most to bind the story tags (in this case two artists) and would therefore highlight something such as an art movement that connected the two artists. The importance of an artwork is contestable and harder to automate, however the approach to theme ranking would prioritize artworks that appeared in the most events. Therefore, an artwork that had appeared in many exhibitions, or featured in other types of event, would be higher ranked. It would though remain the task of the Storyscope user to determine whether this was an accurate indicator of value.

6. CONCLUSIONS AND FUTURE WORK

This work has proposed the use of facts and a simple event structure to make online knowledge sources more accessible to the users of stories. The narrative principles of theme and setting have been applied to facilitate a form of navigation across stories and concepts nearer to the needs of the story author or reader. The Storyscope ontology and RDF export mechanism facilitates exploration across the content of Storyscope sites and the comparison of alternative interpretations. Evaluation suggests the approach has potential in supporting semantic story navigation.

In their analysis of current virtual museum offerings, Geser and Niccolucci [25] argue that automated approaches to creating event-based timelines are unconvincing because the historical narrative and meaningful relationships between the exhibits are lost. We suggest that Storyscope offers a potential solution to this shortcoming in two ways. First, Storyscope can use the set of tags, associated for example with an exhibit or collection of exhibits, to derive the set of themes that bind those tags together and potentially indicate the most meaningful relationships between the tags. Second, Storyscope uses settings, for example associated with the creation of an artwork, to identify larger scale social or political events that constitute a broader historical context.

The approach offered by Storyscope could be applied not only in a virtual museum context for accessing resources outside the museum visit, but also as a method for augmenting the physical museum visit with additional resources. Warren et al [26] describe how the museum experience could be augmented by mapping physical exhibits to additional information (such as a biography of the artist) that could be presented to the visitor on a location-sensitive mobile device. Storyscope could offer a way of automating many of these mappings by deriving additional information related to semantic tags associated with the exhibit. Our recent work using Storyscope technology to augment a sculpture trail in the grounds of a museum [27] suggests this could provide an effective approach to museum visit augmentation.

Given that many museums already represent cultural knowledge using the event-based CIDOC CRM ontology it would be reasonably straightforward to map events from the museum's knowledgebase to the event schema used in Storyscope. Similarly, knowledge aggregated in Storyscope from external sources could be mapped back into facts and events within the museum knowledgebase.

Our immediate plans for development of the Storyscope technology include the following. First, we intend to conduct a fine-grained user study to observe in more detail the cognitive processes involved in story interpretation and authoring and how

they can be supported by Storyscope and the principles of theme and setting. Second, work is underway to extend the range of online data sources used to construct facts and events, including textual resources. Third, we intend to look at whether graph-based approaches and the weighting of event properties could be used to improve the ranking of settings and themes. Fourth, we will look to build on the Storyscope ontology to provide support for exploration across Storyscope sites and the interpretations revealed by different user communities.

7. REFERENCES

- [1] Rowe, S., Wertsch, J., Tatyana, K. Linking Little Narratives to Big Ones: Narrative and Public Memory in History Museums. *Culture and Psychology*, 16 (2): 96-112. 2002.
- [2] <http://www.freebase.com>
- [3] <http://www.mpi-inf.mpg.de/yago-naga/yago>
- [4] <http://factforge.net>
- [5] <http://sig.ma>
- [6] <http://sindice.com>
- [7] Alani, H., Kim, S., Millard, D. E., Weal, M. J., Hall, W., Lewis, P. H. and Shadbolt, N. R. Automatic ontology-based knowledge extraction from web documents. *IEEE Intelligent Systems*, 18(1): 14-21. 2003.
- [8] Wolff, A., Mulholland, P and Collins, T. Storyscope: Using Theme and Setting to Guide Story Enrichment from External Data Sources. In *Proceedings of 24th ACM Conference on Hypertext and Social Media*, Paris, France, pp. 79-88. 2013.
- [9] Polkinghorne, D. *Narrative knowing and the human sciences*. State University of New York Press. 1988.
- [10] Gervás, P. and León, C. The Need for Multi-Aspectual Representation of Narratives in Modelling their Creative Process. In *Proceedings of International Workshop on Computational Models of Narrative*, Quebec City, Canada. 2014.
- [11] Tomashevsky, B. Thematics. In Lemon L. T. and Reis, M. J. ed. *Russian Formalist Criticism*. University of Nebraska Press. pp. 62-95. 1965.
- [12] Le Boeuf, P., Doerr, M., Ore, C. E., Stead, S. eds. *Definition of the CIDOC Conceptual Reference Model*. <http://www.cidoc-crm.org>. 2014.
- [13] Shaw, R., Troncy, R. and Hardman, L. LOD: Linking Open Descriptions of Events. In *Proceedings of Asian Semantic Web Conference*, pp. 153-167. 2009.
- [14] van Hage, W. R., Malaise, V., Segers, R., Hollink, L. and Schreiber, G. Design and use of the Simple Event Model (SEM). *Journal of Web Semantics*, 9(2). 2011.
- [15] Hyvönen, E., Lindquist, T., Törnroos, J. and Mäkelä, E.: History on the Semantic Web as Linked Data - An Event Gazetteer and Timeline for World War I. In *Proceedings of CIDOC*. 2012.
- [16] van den Akker, C., Legêne, S., van Erp, M., Aroyo, L., Segers, R., van der Meij, L., van Ossenbruggen, J., Schreiber, G., Weling, B., Oomen, J. and Jacobs, G. Digital hermeneutics: Agora and the online understanding of cultural heritage. In *Proceedings of ACM Web Science Conference*. 2011.
- [17] <http://www.drupal.org>

- [18] <http://www.alchemyapi.com>
- [19] Mäkelä, E., Hypén, K. and Hyvönen, E. BookSampo - Lessons Learned in Creating a Semantic Portal for Fiction Literature. In *Proceedings of International Semantic Web Conference*. 2011.
- [20] Page, L., Brin, S. and Motwani, R. and Winograd, T. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report. Stanford InfoLab. 1999.
- [21] <http://decipher.open.ac.uk/storyscope/introduction>
- [22] Breslin, J. G. Decker, S. Harth, A. and Bojars, U. SIOC: an approach to connect web-based communities, *International Journal of Web Based Communities*, 2(2006): 133-142.
- [23] Lohmann, S., Díaz, P. and Aedo. I. MUTO: the modular unified tagging ontology, In *Proceedings of International Conference on Semantic Systems*, pp. 95-104. 2011.
- [24] Gangemi A. and Mika P. Understanding the Semantic Web through Descriptions and Situations. In *Proceedings of ODBASE03 Conference*. 2003.
- [25] Geser, G. and Niccolucci, F. Virtual museums, digital reference collections and e-science environments. *Uncommon Culture*, 3(5/6): 12-37. 2012.
- [26] Warren, R., Evans, D., Chen, M., Farrell, M. and Mayles, D. Data-driven enriched exhibits using augmented reality. In *Proceedings of MW2015: Museums and the Web 2015*. <http://mw2015.museumsandtheweb.com/paper/data-driven-augmented-reality-for-museum-exhibits-and-lost-heritage-sites>.
- [27] Wolff, A., Mulholland, P., Maguire, M. and O'Donovan, D. Mobile technology to support coherent story telling across freely explored outdoor artworks. In *Proceedings of Advances in Computer Entertainment*, Madeira. 2014.